MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, **2015/2016**

HPB3011 –Bioinformatics Algorithms II

(All sections / Groups)

11 MARCH 2016
3:00 – 5:00 PM
(2 hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 4 pages, including this cover page.
2. You are required to attempt four (4) out of five (5) questions. All questions carry equal marks (10).
3. Please write all your answers in the Answer Booklet provided.

Question 1

a) What is the principal of DNA microarray technology based on? Briefly explain the types of Microarray?

[3 marks]

b) What is the main purpose of data pre-processing in DNA microarray analysis? What are the three steps used? Briefly describe them.

[4 marks]

c) A microarray data set can be clustered using various approaches, yielding different results. List at least 2 clustering techniques. How can you decide which clustering results are "correct" (most biologically relevant)?

[3 marks]

Question 2

a) Below is the transition probability for CpG (P^+) island and non—CpG (P^-) island chains. Given the sequence "**ACCGGC**", determine whether the sequence is a CpG island.

[3 marks]

$$P^+ = \begin{bmatrix} & A & C & G & T \\ A & 0.18 & 0.27 & 0.43 & 0.12 \\ C & 0.17 & 0.37 & 0.27 & 0.19 \\ G & 0.16 & 0.34 & 0.37 & 0.13 \\ T & 0.08 & 0.36 & 0.38 & 0.18 \end{bmatrix} \quad P^- = \begin{bmatrix} & A & C & G & T \\ A & 0.30 & 0.20 & 0.29 & 0.21 \\ C & 0.32 & 0.30 & 0.08 & 0.30 \\ G & 0.25 & 0.25 & 0.29 & 0.21 \\ T & 0.18 & 0.24 & 0.29 & 0.29 \end{bmatrix}$$

c) In a Support Vector Machine (SVM), what does the equation $y = \mathbf{w}.\mathbf{x} + b$ describe? What does it mean if the value is greater than zero, or if it is less than zero? Why does a SVM have to be trained? How do sensitivity and specificity apply to a SVM?

[3 marks]

d) Use K-means clustering to cluster the data into 2 clusters. For initialization, choose any two points as your centroids. Solve it manually.

Continued...

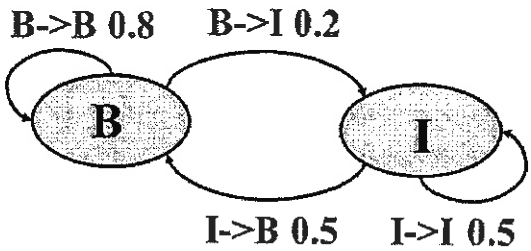
	Gene1	Gene2	Gene3	Gene4	Gene5
sample1	10	11	13	15	18
sample2	9	12	10	17	19

[4 marks]

Question 3

a) Given the following Hidden Markov Model, find the probability of the nucleotide sequence ATT using the forward algorithm.

[6 marks]



P(A|B) = 0.25 P(T|B) = 0.40 P(C|B) = 0.10 P(G|B) = 0.25
P(A|I) = 0.25 P(T|I) = 0.25 P(C|I) = 0.25 P(G|I) = 0.25

b) What are the criteria for Alpha and turns prediction in the Chou - Fasman method?

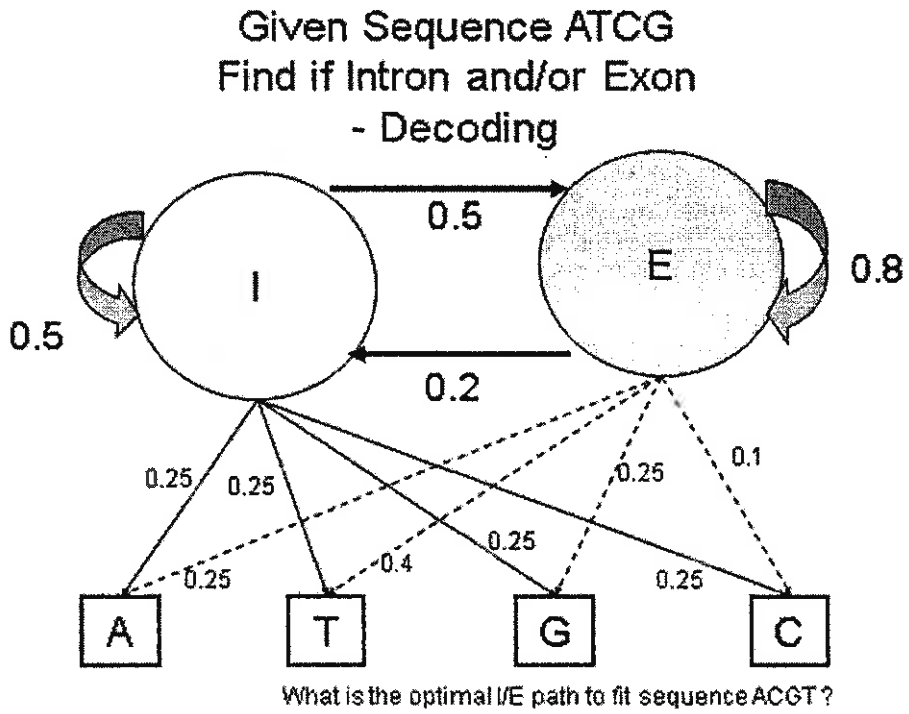
[4 marks]

Question 4

a) Given the sequence ACGT, determine whether it is likely to be intron or exon?

Intron – Exon transition diagram is given below:

Continued...



for Intron P(ATCG) = 0.25 each (emissions or observations)
for Exon P(AG) = 0.25 each, P(C) = 0.1, P(T) = 0.4

find (**decode**) the optimal Intron-Exon path for the sequence ACGT

[6 marks]

b) Briefly describe the principles behind statistical algorithms, nearest neighbour algorithms, and artificial neural networks in protein structure predictions.

[4 marks]

Question 5 [10 marks]

a) What are the steps involved in a protein threading method.

[4 marks]

b) Briefly explain the role of RNA chaperones.

[2 marks]

c) List down 8 different types of RNA and their functions.

[4 marks]

End of Paper.